Auto Process Monroe County Covid-19 Case Data

John Minter

Started 2020-03-30, Last modified: 2020-04-07

Contents

Set up the analysis. Use the `bVerbose` variable as a flag to decide whether or not to print diagnostic data. The default is `FALSE`.

```r
bVerbose <- FALSE

if(bVerbose){
    print(getwd())
}
```

```r
temp <- tempfile()
download.file("https://github.com/nytimes/covid-19-data/archive/master.zip",temp)
us_counties <- read_csv(unz(temp, "covid-19-data-master/us-counties.csv"))
```

```
Parsed with column specification:
cols(
  date = col_date(format = ""),
  county = col_character(),
  state = col_character(),
  fips = col_character(),
  cases = col_double(),
  deaths = col_double()
)
```

```r
path <- paste0(here(), "/us-counties.csv")
write_csv(us_counties, path, append = FALSE, col_names=TRUE)
us_states <- read_csv(unz(temp, "covid-19-data-master/us-states.csv"))
```

```
Parsed with column specification:
cols(
  date = col_date(format = ""),
  state = col_character(),
  fips = col_character(),
  cases = col_double(),
  deaths = col_double()
)
```

```r
pa_states <- paste0(here(), "/us-states.csv")
write_csv(us_states, pa_states, append = FALSE, col_names=TRUE)
unlink(temp)
tail(us_counties)
```

```
# A tibble: 6 x 6
  date       county      state   fips  cases deaths
```

```
      <date>        <chr>       <chr>     <chr> <dbl>   <dbl>
1 2020-04-07 Sheridan     Wyoming 56033    12       0
2 2020-04-07 Sublette     Wyoming 56035     1       0
3 2020-04-07 Sweetwater Wyoming 56037      6       0
4 2020-04-07 Teton        Wyoming 56039    44       0
5 2020-04-07 Uinta        Wyoming 56041     3       0
6 2020-04-07 Washakie     Wyoming 56043     4       0
```

```
tail(us_states)
```

```
# A tibble: 6 x 5
  date        state           fips  cases deaths
  <date>      <chr>           <chr> <dbl>  <dbl>
1 2020-04-07 Virgin Islands 78       45      1
2 2020-04-07 Virginia        51     3333     69
3 2020-04-07 Washington      53     8682    409
4 2020-04-07 West Virginia   54      412      4
5 2020-04-07 Wisconsin       55     2578     94
6 2020-04-07 Wyoming         56      221      0
```

Next, load the data file and extract what we need using functions from the **dplyr** package to create a **tibble** of values (an enhanced R dataframe that works nicely with the **tidyverse** collection of R packages by Hadley Wickham.) We will use the `kable` function from the `knitr` package to get a nice looking table. We really only want the last few values...

```r
pa_cty <- paste0(here(), "/us-counties.csv")
df <- read.csv(pa_cty, header = TRUE, sep = ",")
df$date <- format(as.Date(df$date), "%m-%d")
tib <- as_tibble(df)
tib$date <- as.Date(df$date, "%m-%d")
ny <- tib %>% filter(state == "New York")
monroe_cty <- ny %>% filter(county == "Monroe")
monroe_cty %>% select(date, cases, deaths) -> mc_vals
tail(mc_vals)
```

```
# A tibble: 6 x 3
  date       cases deaths
  <date>     <int>  <int>
1 2020-04-02   420     10
2 2020-04-03   464     14
3 2020-04-04   512     17
4 2020-04-05   548     19
5 2020-04-06   574     23
6 2020-04-07   596     31
```

```
tail(mc_vals)
```

```
# A tibble: 6 x 3
  date       cases deaths
  <date>     <int>  <int>
1 2020-04-02   420     10
2 2020-04-03   464     14
3 2020-04-04   512     17
4 2020-04-05   548     19
5 2020-04-06   574     23
6 2020-04-07   596     31
```

```
print(mc_vals)
```

```
# A tibble: 28 x 3
   date        cases deaths
   <date>      <int> <int>
 1 2020-03-11      1     0
 2 2020-03-12      1     0
 3 2020-03-13      1     0
 4 2020-03-14      2     0
 5 2020-03-15      2     0
 6 2020-03-16     10     0
 7 2020-03-17     11     0
 8 2020-03-18     14     0
 9 2020-03-19     28     0
10 2020-03-20     32     0
# ... with 18 more rows
```

Next, we compute the total deaths for Monroe County

```
tot_deaths <- sum(monroe_cty$deaths)
tot_cases <- sum(monroe_cty$cases)

death_rate_pct <- 100*tot_deaths/tot_cases

death_rate_pct <- round(death_rate_pct, digits = 2)
```

Monroe County's COVID-19 has 5216 cases and 156 deaths with a death rate of 2.99 percent.

Next, we plot the curve.

```
mc_tot_plt <- ggplot(mc_vals, aes(x=date, y=cases)) +
              geom_line(colour='darkred', size=2) +
              xlab("Date") +
              ylab("Total Cases") +
              ggtitle("Monroe County COVID-19 Cases") +
              labs(caption = 'Data from Johns Hopkins') +
              scale_x_date() +
              theme(axis.text=element_text(size=8),
                    axis.title=element_text(size=12),
                    plot.title=element_text(hjust = 0.5)) +
              NULL

print(mc_tot_plt)
```

Monroe County COVID−19 Cases
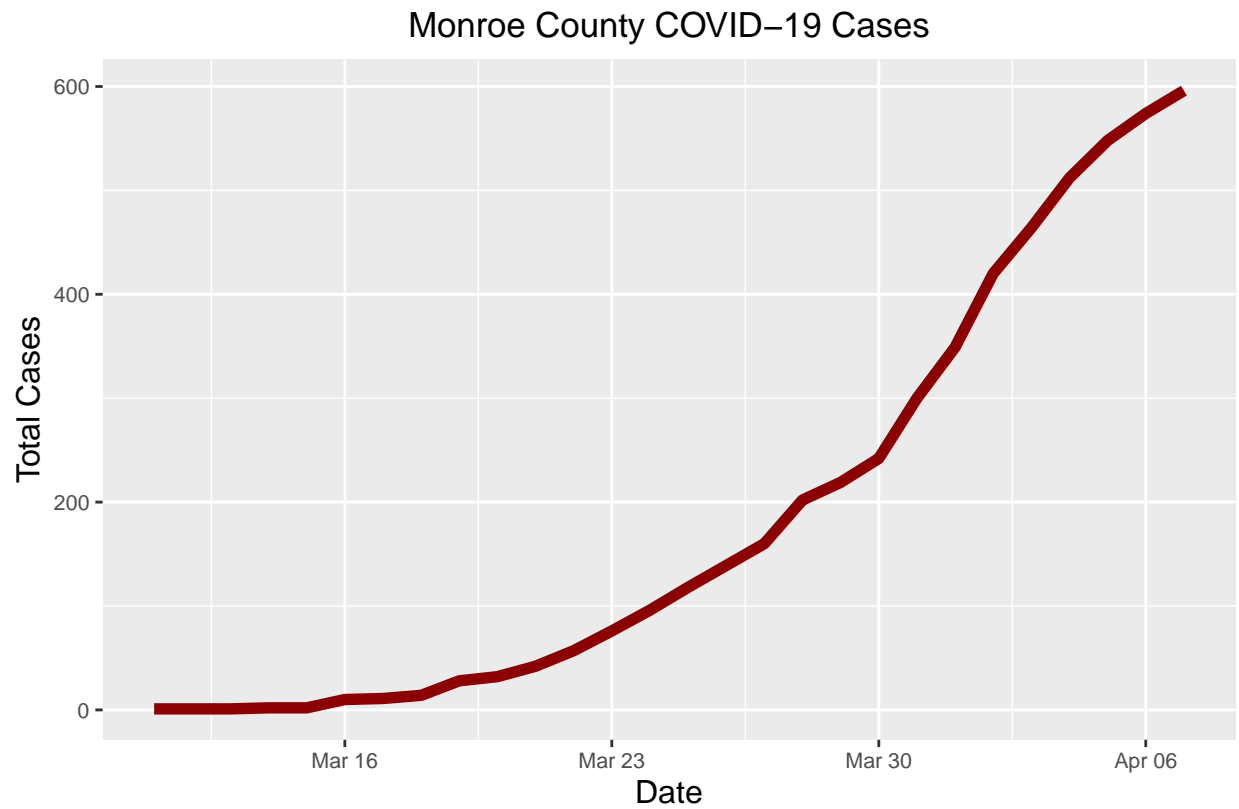
Data from Johns Hopkins

It looks like Monroe County's curve is starting to flatten.

We will save the plot as in both the `png` and `jpg` formats.

```r
strOutPng <- "plt/Monroe_Cty_Covid-19.png"

# we want the png to be close to 1024x768...
ggsave(mc_tot_plt, file=strOutPng, width=9.0, height=6.0,
       units="in", dpi=113.7778)

strOutJpg <- "plt/Monroe_Cty_Covid-19.jpg"

# we want the png to be close to 1024x768...
ggsave(mc_tot_plt, file=strOutJpg, width=9.0, height=6.0,
       units="in", dpi=113.7778)
```